

## MIT Open Access Articles

*Lyapunov Exponent of Rank-One Matrices: Ergodic Formula and Inapproximability of the Optimal Distribution*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Altschuler, Jason M and Parrilo, Pablo A. 2020. "Lyapunov Exponent of Rank-One Matrices: Ergodic Formula and Inapproximability of the Optimal Distribution." SIAM Journal on Control and Optimization, 58 (1).

**As Published:** 10.1137/19M1264072

**Publisher:** Society for Industrial & Applied Mathematics (SIAM)

**Persistent URL:** <https://hdl.handle.net/1721.1/143833>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



# LYAPUNOV EXPONENT OF RANK-ONE MATRICES: ERGODIC FORMULA AND INAPPROXIMABILITY OF THE OPTIMAL DISTRIBUTION\*

JASON M. ALTSCHULER<sup>†</sup> AND PABLO A. PARRILO<sup>†</sup>

**Abstract.** The *Lyapunov exponent* corresponding to a set of square matrices  $\mathcal{A} = \{A_1, \dots, A_n\}$  and a probability distribution  $p$  over  $\{1, \dots, n\}$  is  $\lambda(\mathcal{A}, p) := \lim_{k \rightarrow \infty} \frac{1}{k} \mathbb{E} \log \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\|$ , where  $\sigma_i$  are independent and identically distributed according to  $p$ . This quantity is of fundamental importance to control theory since it determines the asymptotic convergence rate  $e^{\lambda(\mathcal{A}, p)}$  of the stochastic linear dynamical system  $x_{k+1} = A_{\sigma_k} x_k$ . This paper investigates the following “design problem”: Given  $\mathcal{A}$ , compute the distribution  $p$  minimizing  $\lambda(\mathcal{A}, p)$ . Our main result is that it is NP-hard to decide whether there exists a distribution  $p$  for which  $\lambda(\mathcal{A}, p) < 0$ , i.e., it is NP-hard to decide whether this dynamical system can be stabilized. This hardness result holds even in the “simple” case where  $\mathcal{A}$  contains only rank-one matrices. Somewhat surprisingly, this is in stark contrast to the Joint Spectral Radius—the deterministic counterpart of the Lyapunov exponent—for which the analogous optimization problem over switching rules is known to be exactly computable in polynomial time for rank-one matrices. To prove this hardness result, we first observe via Birkhoff’s Ergodic Theorem that the Lyapunov exponent of rank-one matrices admits a simple formula and in fact is a quadratic form in  $p$ . Hardness of the design problem is shown through a reduction from the Independent Set problem. Along the way, simple examples are given illustrating that  $p \mapsto \lambda(\mathcal{A}, p)$  is neither convex nor concave in general. We conclude with extensions to continuous distributions, exchangeable processes, Markov processes, and stationary ergodic processes.

**Key words.** Lyapunov exponent, stochastic linear dynamical system, asymptotic stability, hardness of approximation

**AMS subject classifications.** 93E15, 37M25, 68Q17

**DOI.** 10.1137/19M1264072

**1. Introduction.** Given a finite collection of square matrices  $\mathcal{A} = \{A_1, \dots, A_n\}$  and a probability distribution  $p$  over  $\{1, \dots, n\}$ , the corresponding *Lyapunov exponent*  $\lambda(\mathcal{A}, p)$  is defined as

$$(1.1) \quad \lambda(\mathcal{A}, p) := \lim_{k \rightarrow \infty} \frac{1}{k} \mathbb{E} \log \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\|,$$

where the  $\sigma_i$  are independent and identically distributed (i.i.d.) according to  $p$ . Over the past half-century, this quantity  $\lambda(\mathcal{A}, p)$  has received significant interest due to its many connections and applications to diverse fields including probability, ergodic theory, dynamical systems, functional analysis, representation theory, computer image generation of fractals, control theory, and more; see, e.g., the surveys [4, 9, 10, 16, 21, 47] and the references therein.

The primary application of the Lyapunov exponent that this paper focuses on is in control theory. The fundamental connection is that  $\lambda(\mathcal{A}, p)$  dictates the asymptotic

\*Received by the editors May 28, 2019; accepted for publication December 17, 2019; published electronically February 25, 2020. A preliminary version of this work appeared at the 2019 Conference on Decision and Control. The present paper is a significant extension of that work, and in particular all results in section 5 here are new. This includes extensions to continuous distributions, exchangeable processes, Markov processes, and stationary ergodic processes.

<https://doi.org/10.1137/19M1264072>

**Funding:** This work was partially supported by NSF AF 1565235 and NSF Graduate Research Fellowship 1122374.

<sup>†</sup>Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139 (jasonalt@mit.edu, parrilo@mit.edu).

growth rate of the stochastic linear dynamical system

$$(1.2) \quad x_{k+1} = A_{\sigma_k} x_k.$$

Indeed, the celebrated Furstenberg–Kesten theorem [23] establishes that (under a mild integrability condition) the average norm  $\lim_{k \rightarrow \infty} \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\|^{1/k}$  almost surely (a.s.) exists and is equal to the *Lyapunov spectral radius*

$$(1.3) \quad R(\mathcal{A}, p) := e^{\lambda(\mathcal{A}, p)}.$$

In particular, the dynamical system (1.2) converges asymptotically<sup>1</sup> to zero (a.s.) if and only if  $R(\mathcal{A}, p) < 1$ , i.e., if and only if  $\lambda(\mathcal{A}, p) < 0$ . Note that in this sense, the Lyapunov spectral radius is a natural stochastic generalization of the spectral radius of a single matrix. Two natural control-theoretic questions arise about the dynamical system (1.2): analysis and design.

The *analysis problem* is to compute the convergence rate  $\lambda(\mathcal{A}, p)$  of the dynamical system given (fixed) parameters  $(\mathcal{A}, p)$ . This problem is well known to be difficult, and approximating  $\lambda(\mathcal{A}, p)$  is in general algorithmically undecidable [46, Theorem 2]. Nevertheless, the significance of this problem has motivated a substantial body of research on algorithms for obtaining upper or lower bounds on  $\lambda(\mathcal{A}, p)$ , and on identifying special cases in which this problem becomes tractable; see, e.g., [6, 13, 14, 17, 18, 24, 30, 32, 34, 40, 41, 44].

The *design problem*—which is the focus of this paper—is to optimize the dynamical system so that the convergence is as fast as possible, namely, given a fixed set of matrices  $\mathcal{A}$ , compute a distribution  $p$  minimizing  $\lambda(\mathcal{A}, p)$ . In contrast to the analysis question, there seems to be little previous work on this design problem. (A related version of this design problem uses instead the more restrictive notion of convergence in *mean*; see, e.g., [8].) One of this paper’s primary contributions is to explain this lack of progress on the design problem by showing that it is NP-hard to determine whether there exists a distribution  $p$  such that  $\lambda(\mathcal{A}, p) < 0$ . That is, unless  $P = NP$ , there is no polynomial-time algorithm for deciding whether the dynamical system (1.2) has a stabilizing distribution  $p$ . Moreover, this complexity result holds even in the “simple” case where all matrices in  $\mathcal{A}$  are rank one.

**1.1. Contributions and outline.** This paper focuses on the case when all matrices in  $\mathcal{A}$  are *rank one*. The advantage of restricting to this simple case is that (i) the analysis problem becomes tractable as  $\lambda(\mathcal{A}, p)$  admits a simple, closed-form expression, and (ii) this formula enables us to prove that the design problem is hard in a complexity-theoretic sense. We elaborate below on details and implications. Table 1 contains an overview of our main results.

Section 2 studies the analysis problem when  $\mathcal{A}$  is a collection of rank-one matrices. We observe in Theorem 2.1 that in this special case,  $\lambda(\mathcal{A}, p)$  admits a simple closed-form expression and is in fact a *quadratic form* in  $p$ . This is an ergodic-type result since it equates  $\lambda(\mathcal{A}, p)$ —which is a “time average” of the product of an infinite sequence of random matrices—to a simpler “space average.” We note that this formula readily

<sup>1</sup>The Lyapunov exponent characterizes a.s. convergence of (1.2). In general, this is different from convergence in *mean*, i.e.,  $\lim_{k \rightarrow \infty} \mathbb{E} \|A_{\sigma_k} \cdots A_{\sigma_1}\| = 0$ . For example, consider the scalar case with  $n = 2$ , where  $A_{\sigma_k}$  are i.i.d., and are either 2 or  $\frac{1}{100}$  with probability  $\frac{1}{2}$ . Then  $A_{\sigma_k} \cdots A_{\sigma_1}$  does not converge in mean (in fact  $\mathbb{E} |A_{\sigma_k} \cdots A_{\sigma_1}|$  grows exponentially in  $k$ ) due to a few exponentially large but exponentially unlikely trajectories, but *does* converge to 0 with probability 1 (since  $\exp(\mathbb{E} \log |A_{\sigma_1}|) < 1$ ).

TABLE 1

*Complexity overview of the analysis and design questions for the Lyapunov exponent—for collections of general matrices, as well as the special case of rank-one matrices. The various versions of the design problem are for optimizing over different classes of stochastic processes. Our new results are shown in italics.*

	Analysis question	Design question (i.i.d.)	Design question (exchangeable)	Design question (Markov)
Rank one	<i>Poly-time alg.</i> (Thm. 2.1)	<i>NP-hard</i> (Thm. 4.1)	<i>NP-hard</i> (Thm. 5.3)	<i>Poly-time alg.</i> (Thm. 5.4)
General	Undecidable [46, Thm. 2]	<i>NP-hard</i> (Thm. 4.1)	<i>NP-hard</i> (Thm. 5.3)	?

extends to ergodic stationary processes taking values in sets  $\mathcal{A}$  consisting of infinitely many rank-one matrices; however, for expository reasons we defer this generalization to section 5, and provide in section 2 just the simple case of i.i.d. processes taking values in finite sets  $\mathcal{A}$ , since this admits an elementary proof and requires essentially no measure theory. We remark that this formula shows that for this rank-one case,  $\lambda(\mathcal{A}, p)$  is computable in polynomial time (as opposed to being undecidably hard to approximate in the general case [46]).

Sections 3 and 4 then turn to the design problem. Specifically, section 3 investigates the convexity and concavity of the function  $p \mapsto \lambda(\mathcal{A}, p)$  over the probability simplex  $\Delta_n$ . By our aforementioned result that  $\lambda(\mathcal{A}, p)$  is a quadratic form in this rank-one case, these convexity/concavity properties are equivalent to the conditional definiteness of the Hessian  $\frac{\partial^2}{\partial p^2} \lambda(\mathcal{A}, p)$ . We give simple, small examples illustrating that in general this matrix is neither conditionally positive semidefinite nor negative semidefinite, and then explain this phenomenon by making a connection to the fact that the Martin distance on the  $(1, n)$  Grassmannian is not a metric.

Of course, this lack of convexity of the function  $p \mapsto \lambda(\mathcal{A}, p)$  does not in itself prove that the design problem  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  is hard. Section 4 formally shows that it is NP-hard to even determine the sign of  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$ . That is, unless  $P = NP$ , there is no polynomial-time algorithm for determining whether the dynamical system (1.2) has a stabilizing distribution  $p$ . This explains the lack of prior algorithmic progress on the design problem. The key idea in the proof is a reduction from the Independent Set problem, which is known to be NP-hard to approximate and is also known to be expressible as the minimization of a certain quadratic form over the probability simplex. The intuition for this reduction is, roughly speaking, that the optimal distribution for  $\lambda(\mathcal{A}, p)$  should concentrate mass on a set of matrices in  $\mathcal{A}$  which are as mutually “orthogonal” to each other as possible, and such a set of matrices can be thought of as an independent set in a certain associated graph. (See the proof of Theorem 4.1 for precise details.)

These hardness results uncover an interesting and somewhat surprising (at least to the authors) dichotomy between the complexity of optimizing the Lyapunov exponent and its deterministic analogue, the *Joint Spectral Radius* (JSR). Recall that the JSR of a collection of square matrices  $\mathcal{A} = \{A_1, \dots, A_n\}$  is defined as

$$(1.4) \quad \rho(\mathcal{A}) := \lim_{k \rightarrow \infty} \max_{\sigma \in \{1, \dots, n\}^k} \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\|^{1/k}.$$

The JSR is a natural generalization of the spectral radius of a single matrix and characterizes the maximal (i.e., worst-case) asymptotic growth rate of the linear dynamical system (1.2) over all possible choices of switching rules  $\sigma \in \{1, \dots, n\}^k$ . Due to its many applications (see, e.g., [29] and the references within), the JSR has also

attracted a large body of research. Similarly to the Lyapunov exponent, the JSR is known to in general have algorithmic undecidability and NP-hardness complexity barriers [7, 46], yet nevertheless there has been significant work on approximation algorithms for  $\rho(\mathcal{A})$  and on identifying special cases in which the problem becomes more tractable [1, 26, 29, 38, 42]. A specific such setting that has received significant attention is when the matrices are rank one; see, e.g., [2, 27, 33, 35, 36, 39]. In particular, it is known that for rank-one matrices, the optimization problem over switching rules  $\sigma \in \{1, \dots, n\}^k$  required to compute the JSR is equivalent to a certain discrete optimization problem over an associated graph, namely the Maximum-Mean-Cycle problem, which can be solved in polynomial time [2, 27, 31]. This is in stark contrast to our result that the Lyapunov exponent of rank-one matrices is NP-hard to optimize, even approximately.

In subsection 4.2, we note that in the special case when the set  $\mathcal{A}$  is “well-conditioned,” there is a polynomial-time approximation algorithm for  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$ .

Finally, section 5 considers extensions to more general stochastic processes  $A_{\sigma_1}, A_{\sigma_2}, \dots$ . The corresponding analysis and design problems are discussed in subsections 5.1 and 5.2, respectively. Specifically, subsection 5.1 shows that a similar formula holds for the “Lyapunov exponent” of ergodic stationary processes of rank-one matrices. This allows for the study of more exotic stochastic linear dynamical systems (1.2). Using this formula, we compute the convergence rate for setups including i.i.d. processes over continuous distributions, exchangeable processes, and Markov processes. As an example application, we explicitly calculate the Lyapunov exponent for the induced distribution over matrices  $uu^T$ , where  $u$  is distributed according to the  $d$ -dimensional spherical measure. Subsection 5.2 then turns to the design problem for these more general setups. We show that optimizing the convergence rate of the dynamical system (1.2) is NP-hard over exchangeable processes, but can be done in polynomial time for Markov processes. The intuition for these results is as follows. First, we observe that (after some manipulation) these problems seek to optimize a linear functional over a convex set, and thus are equivalent to optimizing over the extreme points. Informally speaking, the extreme points of the sets of exchangeable processes and Markov processes, respectively, are the sets of i.i.d. processes and deterministic processes. The arguments are then finished by recalling that optimizing the convergence rate over i.i.d. processes is NP-hard (discussed above), while optimizing over deterministic processes is known to be polynomial-time computable.

**1.2. Notation.** The set of discrete probability distributions on  $n$  atoms is identified with the simplex  $\Delta_n := \{p \in \mathbb{R}_{\geq 0}^n : \sum_{i=1}^n p_i = 1\}$ . The support of a discrete probability distribution  $p \in \Delta_n$  is denoted by  $\text{supp}(p) := \{i \in [n] : p_i > 0\}$ . We write  $X \sim p$  to denote that  $X$  is a random variable with distribution  $p$ . The symbols  $e_i$ ,  $\mathbf{1}$ , and  $I$  denote the  $i$ th standard basis vector, all-ones vector, and identity matrix, respectively, in an ambient dimension that will be clear from the context. The norm  $\|\cdot\|_2$  denotes the Euclidean norm when applied to vectors, and denotes the spectral norm when applied to matrices. All logarithms are natural logarithms, i.e., taking base  $e$ . We abbreviate “almost surely” by a.s., “without loss of generality” by w.l.o.g., “independent and identically distributed” by i.i.d., “positive semidefinite” by PSD, and “negative semidefinite” by NSD. Recall that a square matrix  $M$  is said to be conditionally PSD (resp., conditionally NSD) if  $p^T M p \geq 0$  (resp.,  $\leq 0$ ) for all vectors  $p$  in the orthogonal component of the subspace spanned by  $\mathbf{1}$ . All other specific notation is introduced in the main text.

**2. Ergodic formula for Lyapunov exponent of rank-one matrices.** This section gives an explicit closed-form expression for the Lyapunov exponent of a collection of rank-one matrices. This is an ergodic-type result in the sense that it equates the Lyapunov exponent—which is a “time average”—to a simpler “space average.”

This result readily extends to stationary ergodic processes where the collection  $\mathcal{A}$  contains infinitely many matrices (Theorem 5.1). However, for simplicity of exposition, we first present the discrete case (Theorem 2.1) since it admits an elementary proof with essentially no measure theory.

**THEOREM 2.1.** *For any collection of rank-one square matrices  $\mathcal{A} := \{A_i = u_i v_i^T\}_{i=1}^n$  and any probability distribution  $p \in \Delta_n$ ,*

$$(2.1) \quad \lambda(\mathcal{A}, p) = \sum_{i,j=1}^n p_i p_j \log |u_i^T v_j|.$$

*Remark 2.2.* The formula in Theorem 2.1 is independent of the rank-one decompositions  $A_i = u_i v_i^T$ . Indeed, a rescaling of  $u_i$  and  $v_i$  changes the cost matrix  $[\log |u_i^T v_j|]_{i,j}$  by a skew-symmetric matrix, which does not affect the underlying quadratic form  $\sum_{i,j} p_i p_j \log |u_i^T v_j|$ .

It is important to emphasize that, in general, the Lyapunov exponent of a set of matrices *cannot* be computed in closed form. It is well known that (under mild conditions) the Lyapunov exponent is equal to a simpler “space average” over a certain invariant distribution [22]; however, the complexity of such ergodic formulas lies in the invariant distribution, which is often difficult—or impossible—to characterize or manipulate analytically. The key difference for the rank-one setting studied here is that in this special case, the invariant distribution has a simple closed form. This is because in this rank-one case, the contraction of the next matrix  $A_{\sigma_{k+1}}$  depends on the previous matrices essentially only through the last matrix  $A_{\sigma_k}$ .

*Proof of Theorem 2.1.* Note that the definition of  $\lambda(\mathcal{A}, p)$  is independent of the norm  $\|\cdot\|$ ; for convenience of computation, let us take it to be the spectral norm. Let  $\sigma_1, \sigma_2, \dots$  be i.i.d. according to  $p$ , and expand the (random) quantity

$$(2.2) \quad \begin{aligned} \frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\|_2 &= \frac{1}{k} \log \|(u_{\sigma_k} v_{\sigma_k}^T) \cdots (u_{\sigma_2} v_{\sigma_2}^T)(u_{\sigma_1} v_{\sigma_1}^T)\|_2 \\ &= \frac{1}{k} \sum_{t=1}^{k-1} \log |u_{\sigma_t}^T v_{\sigma_{t+1}}| + \frac{1}{k} (\log \|u_{\sigma_k}\|_2 + \log \|v_{\sigma_1}\|_2). \end{aligned}$$

We may assume w.l.o.g. that  $A_i \neq 0$  for all  $i \in \text{supp}(p)$ , since otherwise the proof is immediate as both sides of (2.1) equal  $-\infty$ . Thus the latter term  $\frac{1}{k}(\log \|u_{\sigma_k}\|_2 + \log \|v_{\sigma_1}\|_2)$  in (2.2) converges to 0 as  $k \rightarrow \infty$ .

Now let  $G_{\mathcal{A}}$  be the complete directed graph on  $n$  nodes, one for each matrix  $A_i$ , with weight  $\log |u_i^T v_j|$  on the edge from node  $i$  to node  $j$ . Consider the  $n$ -state Markov chain on  $G_{\mathcal{A}}$  with transition probability  $p_j$  from node  $i$  to node  $j$ , irrespective of  $i$ . The key point is that after multiplying by  $\frac{k}{k-1}$ , the sum in (2.2) is precisely the average weight  $\frac{1}{k-1} \sum_{t=1}^{k-1} \log |u_{\sigma_t}^T v_{\sigma_{t+1}}|$  of edges travelled when running this Markov chain for  $k-1$  steps from initial state  $\sigma_1 \sim p$ . Now since this Markov chain has stationary distribution  $p$ , a standard ergodic result on discrete-time, finite-state Markov chains (see, e.g., [25]) shows that this quantity converges a.s. to  $\sum_{i,j} p_i p_j \log |u_i^T v_j|$ . Thus

by (2.2) and the fact that  $\lim_{k \rightarrow \infty} \frac{k-1}{k} = 1$ , we conclude that

$$(2.3) \quad \lim_{k \rightarrow \infty} \frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\| \stackrel{\text{a.s.}}{=} \sum_{ij} p_i p_j \log |u_i^T v_j|.$$

By submultiplicativity of the spectral norm, we have that  $|\frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\|_2| \leq \max_{i \in \text{supp}(p)} |\log \|A_i\|_2|$  is uniformly bounded for all  $k \in \mathbb{N}$ . Thus the claimed formula for  $\lambda(\mathcal{A}, p)$  follows from the a.s. convergence proven in (2.3) and an application of Lebesgue's Dominated Convergence Theorem.  $\square$

*Remark 2.3.* Theorem 2.1 shows that the Lyapunov exponent of rank-one matrices depends only on the products  $A_i A_j$  of length *two*. This is in contrast to the JSR of rank-one matrices, which depends on all matrix products of length up to  $n$  [2, Corollary 2.1].

*Remark 2.4.* Theorem 2.1 shows that the Lyapunov exponent of rank-one matrices is not only computable, but also computable in polynomial time. This is in sharp contrast to the Lyapunov exponent of general matrices (i.e., not necessarily rank-one), which is *undecidably hard* to approximate [46, Theorem 2].

*Remark 2.5.* It is an interesting open question whether such results extend to collections of matrices with some small fixed rank  $k > 1$ . In particular, is there a closed-form expression for the Lyapunov exponent, analogously to Theorem 2.1? And if so, how do the phenomena mentioned in Remarks 2.3 and 2.4 change?

**3. Convexity properties of the Lyapunov exponent as a function of the distribution.** From Theorem 2.1, we know that  $\lambda(\mathcal{A}, p)$  is a quadratic form in  $p$ :

$$(3.1) \quad \lambda(\mathcal{A}, p) = p^T M_{\mathcal{A}} p,$$

where  $(M_{\mathcal{A}})_{ij} := \frac{1}{2}(\log |u_i^T v_j| + \log |u_j^T v_i|)$ . This resolves the *analysis problem* of computing the Lyapunov exponent given a collection of rank-one matrices  $\mathcal{A}$  and a probability distribution  $p$ . In this and the following section, we now turn to the *design problem*: Given a collection of matrices  $\mathcal{A}$ , find the distribution  $p \in \Delta_n$  minimizing  $\lambda(\mathcal{A}, p)$ . As mentioned in the introduction, this is a fundamental problem in control theory with applications including, e.g., fast convergence in the corresponding stochastic linear dynamical system (1.2).

We begin with an example illustrating that even in simple cases, the optimal distribution can depend in a nonobvious way on the geometric configuration of the matrices in  $\mathcal{A}$ .

*Example 3.1.* We consider several cases of symmetric rank-one matrices  $\mathcal{A} = \{u_i u_i^T\}_{i=1}^3$  (plotted in Figure 1) to illustrate the design problem  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$ . This problem has a simple and intuitive geometric interpretation in terms of the linear dynamical system<sup>2</sup>  $x_{k+1} = A_{\sigma_k} x_k = u_{\sigma_k} u_{\sigma_k}^T x_k$ , which corresponds to projecting (and scaling) the current state vector  $x_k$  onto a randomly chosen vector  $u_{\sigma_k}$ . Specifically,  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  corresponds to finding the distribution  $p \in \Delta_3$  over lines  $\{u_i\}_{i=1}^3$  such that repeatedly projecting (and scaling) onto a random vector  $u_i$  with probability  $p_i$  converges as quickly as possible. In all examples,  $u_1 = [1, 10]^T$  and  $u_2 = [10, 1]^T$  are close to orthogonal in that they have small inner product; this is why the optimal distribution is concentrated on them in Figure 1(a). In Figure 1(b),  $u_3$  has smaller

<sup>2</sup>This can be viewed, e.g., as a randomized instance of Dykstra's alternating-projection method for finding a point at the intersection of convex sets (namely the lines  $u_i$  here) [20].

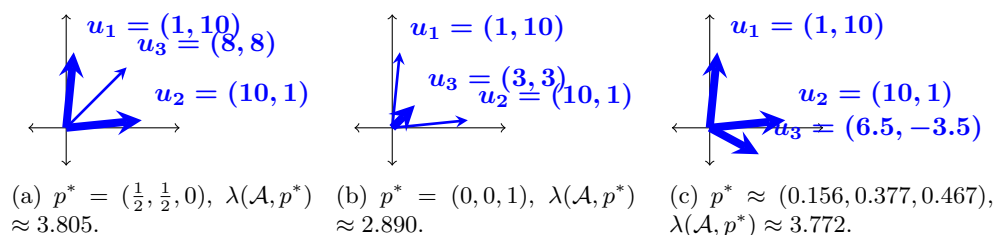


FIG. 1. Vectors  $\{u_i\}_{i=1}^3$  for the three scenarios in Example 3.1. Boldface vectors correspond to the support of the optimal distribution  $p^* := \operatorname{argmin}_{p \in \Delta_3} \lambda(\mathcal{A}, p)$ , where  $\mathcal{A} := \{u_i u_i^T\}_{i=1}^3$ .

norm, and the optimal distribution is concentrated just on it. In Figure 1(c), the optimal distribution is supported on all three matrices. Finally, we note that the optimal distribution for the *maximization* problem  $\max_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  is easily characterized: It is always concentrated on the vector(s) of largest norm (see Lemma 4.8 below).

In the rest of the section, we investigate the (lack of) convexity and concavity of the function  $p \mapsto \lambda(\mathcal{A}, p)$  for a fixed collection of rank-one matrices  $\mathcal{A} = \{u_i v_i^T\}_{i=1}^n$ . Algorithmic and complexity results for the design problem  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  are then discussed in the following section.

Note that since  $p \in \Delta_n$  is a probability distribution in our context, the function  $\Delta_n \ni p \mapsto \lambda(\mathcal{A}, p) = p^T M_{\mathcal{A}} p$  is convex (resp., concave) if and only if the matrix  $M_{\mathcal{A}}$  is conditionally PSD (resp., conditionally NSD). We begin by observing that these conditional semidefiniteness properties of  $M_{\mathcal{A}}$  are invariant under rescaling of the matrices in  $\mathcal{A}$ .

LEMMA 3.2. *The conditional semidefiniteness of  $M_{\mathcal{A}}$  is invariant under nonzero rescalings of matrices in  $\mathcal{A}$ .*

*Proof.* It suffices to show this invariance holds under rescaling  $u_i$  by some  $\alpha \neq 0$ . This scaling changes  $M_{\mathcal{A}}$  to  $M_{\mathcal{A}} + \Delta$  where  $\Delta := \frac{1}{2} \log |\alpha| (e_i \mathbf{1}^T + \mathbf{1} e_i^T)$ . Now for any vector  $x$  orthogonal to  $\mathbf{1}$ ,  $x^T \Delta x = 0$ ; thus in particular  $x^T (M_{\mathcal{A}} + \Delta) x = x^T M_{\mathcal{A}} x$ . Therefore  $M_{\mathcal{A}} + \Delta$  is conditionally PSD (resp., NSD) if and only if  $M_{\mathcal{A}}$  is.  $\square$

As seen next,  $M_{\mathcal{A}}$  is neither conditionally PSD nor conditionally NSD in general. This is true even if all matrices in  $\mathcal{A}$  are symmetric, i.e., when each  $u_i = v_i$ , so the rest of this section henceforth restricts to this symmetric case  $\mathcal{A} = \{u_i u_i^T\}_{i=1}^n$  for simplicity of exposition.

Example 3.3. Let  $\mathcal{A} = \{e_1 e_1^T, e_2 e_2^T\} \subset \mathbb{R}^{2 \times 2}$ , where  $e_1 := [1, 0]^T$  and  $e_2 := [0, 1]^T$ . Then  $M_{\mathcal{A}} = \begin{bmatrix} 0 & -\infty \\ -\infty & 0 \end{bmatrix}$  is not conditionally NSD since  $\begin{bmatrix} 1 & -1 \end{bmatrix} M_{\mathcal{A}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \infty \not\leq 0$ .

Example 3.4. Let  $\mathcal{A} = \{e_1 e_1^T, e_2 e_2^T, (e_1 + e_2)(e_1 + e_2)^T\} \subset \mathbb{R}^{2 \times 2}$ , where  $e_1 := [1, 0]^T$  and  $e_2 := [0, 1]^T$ . Then  $M_{\mathcal{A}} = \begin{bmatrix} 0 & -\infty & 0 \\ -\infty & 0 & 0 \\ 0 & 0 & \log 2 \end{bmatrix}$  is not conditionally PSD since  $\begin{bmatrix} 1 & 1 & -2 \end{bmatrix} M_{\mathcal{A}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = -\infty \not\geq 0$ .

Note that Examples 3.3 and 3.4 are easily modified so that  $M_{\mathcal{A}}$  only has finite entries: add  $\varepsilon \mathbf{1}$  to each of the vectors  $u_i$ , for some small  $\varepsilon > 0$ .

We conclude the section by connecting this lack of convexity of  $p \mapsto \lambda(\mathcal{A}, p)$  to the Martin distance on the space of lines (i.e., the  $(1, n)$  Grassmannian). For simplicity, we continue to restrict to the symmetric case  $A_i = u_i u_i^T$  and assume that the vectors



$u_i$  are normalized to have unit Euclidean norm (this does not change convexity by Lemma 3.2). Since  $-M_{\mathcal{A}}$  has diagonal entries of zero (because of the normalization  $\|u_i\|_2 = 1$ ) and off-diagonal entries that are strictly positive (by Cauchy–Schwarz), it follows from a celebrated result of Schoenberg that  $-M_{\mathcal{A}}$  is conditionally NSD if and only if there is an embedding  $z_1, \dots, z_n \in \ell_2$  such that the Euclidean distance  $\|z_i - z_j\|_2 = \sqrt{-(M_{\mathcal{A}})_{ij}}$  for each  $i, j \in \{1, \dots, n\}$  [43, Theorem 1]. In particular, this implies that if  $M_{\mathcal{A}}$  were conditionally PSD, then  $\{\sqrt{-(M_{\mathcal{A}})_{ij}}\}_{i,j=1}^n$  must be an embeddable collection of Euclidean distances, and thus in particular must satisfy the triangle inequality. However,

$$\sqrt{-(M_{\mathcal{A}})_{ij}} = \sqrt{-\log |u_i^T u_j|}$$

is precisely the Martin distance between the lines in directions  $u_i$  and  $u_j$ , and the Martin distance is well known to *not* satisfy the triangle inequality (i.e., it is not a metric<sup>3</sup>) [15, section 12]. Therefore  $M_{\mathcal{A}}$  cannot always be conditionally PSD.

**4. Optimizing the distribution.** We now investigate the design problem from an algorithmic viewpoint. First we show in subsection 4.1 that this problem is NP-hard in general; then in subsection 4.2, we briefly note that in a certain “well-conditioned” case, there is a polynomial-time approximation algorithm.

**4.1. Inapproximability.** The main result here (Theorem 4.1) is that even the stabilization decision problem for Lyapunov exponents is NP-hard. That is, unless  $P = NP$ , there is no polynomial-time algorithm for deciding whether there exists a distribution  $p \in \Delta_n$  satisfying  $\lambda(\mathcal{A}, p) < 0$ . Moreover, these hardness results hold even in the “simple” case where  $\mathcal{A}$  contains only rank-one symmetric matrices.

**THEOREM 4.1.** *Given a collection  $\mathcal{A}$  of  $n$  symmetric rank-one matrices, it is NP-hard to decide whether*

$$(4.1) \quad \min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$$

*is negative.*

A few remarks are in order. First, note that Theorem 4.1 can be equivalently restated in terms of the Lyapunov spectral radius  $R(\mathcal{A}, p) = e^{\lambda(\mathcal{A}, p)}$  (see (1.3)). That is, it is NP-hard to decide whether

$$(4.2) \quad \min_{p \in \Delta_n} R(\mathcal{A}, p)$$

is less than 1 for a collection  $\mathcal{A}$  of  $n$  symmetric rank-one matrices. Recall from the introduction that  $R(\mathcal{A}, p) < 1$  is equivalent to the (a.s.) asymptotic convergence of the stochastic linear dynamical system (1.2).

Next, we emphasize that this hardness result is somewhat surprising (at least to the authors) in the sense that the analogous optimization problem over switching rules for computing the JSR—which is the deterministic counterpart of the Lyapunov exponent (see the introduction for details)—can be computed exactly in polynomial time for rank-one matrices [2, Remark 2.1].

<sup>3</sup>For example, consider the (rescaled) construction from Example 3.4:  $u_1 = e_1$ ,  $u_2 = e_2$ , and  $u_3 = (e_1 + e_2)/\sqrt{2}$ . Then  $\sqrt{-\log |u_1^T u_3|} + \sqrt{-\log |u_2^T u_3|} = \sqrt{2 \log 2} < \infty = \sqrt{-\log |u_1^T u_2|}$ , which shows that the triangle inequality does not hold for the Martin distance.

We now turn to proving Theorem 4.1. Note that by Theorem 2.1, it suffices to show the NP-hardness of deciding the sign of the following quadratic optimization problem over the simplex:

$$(4.3) \quad \min_{p \in \Delta_n} p^T M_{\mathcal{A}} p,$$

where  $(M_{\mathcal{A}})_{ij} = \log |u_i^T u_j|$ . We observed in section 3 that, in general, this optimization problem (4.3) is not convex since  $M_{\mathcal{A}}$  is not conditionally PSD. Although this lack of convexity is a first step towards suggesting the hardness of the minimization problem (4.3), it of course does not constitute a hardness proof in itself.

We will prove Theorem 4.1 via a reduction from approximating the Independent Set problem. It is helpful to first recall two classical results, namely the Motzkin–Straus formulation for the Independent Set problem as a quadratic optimization over the simplex, and the hardness of approximation of the Independent Set problem. Below,  $\alpha(G)$  denotes the *independence number* of a graph  $G$ , i.e., the size of the largest subset of nodes such that no pair of constituent nodes is connected by an edge.  $A_G$  denotes the adjacency matrix of  $G$ , i.e., the entry  $(A_G)_{ij}$  is 1 if  $(i, j)$  is an edge of  $G$  and is 0 otherwise.

THEOREM 4.2 ([37, Corollary 1]). *For any undirected graph  $G$ , we have  $\frac{1}{\alpha(G)} = \min_{p \in \Delta_n} p^T (I + A_G) p$ .*

THEOREM 4.3 ([48, Theorem 1.1]).<sup>4</sup> *For all  $\varepsilon > 0$ , it is NP-hard to approximate the independence number of a graph on  $n$  nodes to within a multiplicative factor of  $n^{1-\varepsilon}$ .*

The reduction for proving Theorem 4.1 is as follows: Given an undirected graph  $G$  on  $n$  nodes and with edge set  $E$ , let  $B := 3nI + \exp[I + A_G]$ , where the exponential is applied entrywise. Note that  $B$  is symmetric and diagonally dominant since for each  $i \in [n]$ ,  $B_{ii} = 3n + e \geq n + (e - 1) \cdot \text{degree}(i) - 1 = \sum_{j \neq i} B_{ij}$ . Thus  $B$  is PSD and in particular admits a factorization  $B = U^T U$  for some  $U \in \mathbb{R}^{d \times n}$ ,  $d \leq n$ . Denote the columns of  $U$  by  $u_1, \dots, u_n \in \mathbb{R}^d$ , and let  $M \in \mathbb{R}^{n \times n}$  be the matrix with entries  $M_{ij} := \log |u_i^T u_j| = \log |(U^T U)_{ij}| = \log |B_{ij}| = \log B_{ij}$ . Now the key observation is that  $\min_{p \in \Delta_n} p^T M p$  is an  $O(\log n)$  multiplicative approximation to  $\frac{1}{\alpha(G)}$ .

LEMMA 4.4.  $\frac{1}{\alpha(G)} \leq \min_{p \in \Delta_n} p^T M p \leq O(\log n) \frac{1}{\alpha(G)}$ .

*Proof.* Since  $M = (I + A_G) + (\log(3n + e) - 1)I$ , we have

$$(4.4) \quad I + A_G \leq M \leq O(\log n) (I + A_G),$$

where the matrix inequalities are in the entrywise sense. Thus

$$(4.5) \quad \min_{p \in \Delta_n} p^T (I + A_G) p \leq \min_{p \in \Delta_n} p^T M p \leq O(\log n) \min_{p \in \Delta_n} p^T (I + A_G) p.$$

The proof is then complete by applying Theorem 4.2.  $\square$

The last ingredient we will need for the proof of Theorem 4.1 is the simple but helpful observation that  $\lambda(\mathcal{A}, p)$  additively shifts if all matrices in  $\mathcal{A}$  are scaled by the same constant. This follows immediately from the definition of the Lyapunov exponent—or equivalently, by homogeneity of the Lyapunov spectral radius.

<sup>4</sup>The complexity result stated in [48]—as well as in much of the related literature—is for the problem of computing the size of the largest clique in a graph. However, this problem is well known to be equivalent to the Independent Set problem, since a set of nodes is a clique in a graph if and only if it is an independent set in the complement graph.

OBSERVATION 4.5. Let  $\mathcal{A}$  be some collection of square matrices, and let  $c\mathcal{A} := \{cA : A \in \mathcal{A}\}$  for some positive scalar  $c$ . Then  $\lambda(c\mathcal{A}, p) = \lambda(\mathcal{A}, p) + \log c$  for all  $p \in \Delta_n$ .

*Proof of Theorem 4.1.* Assume there exists a polynomial-time algorithm for this decision problem. Now consider an undirected graph  $G$  on  $n$  nodes, and note that the aforementioned matrix  $M$  and vectors  $\{u_i\}_{i=1}^n$  are computable in polynomial time in  $n$ . Letting  $\mathcal{A} := \{u_i u_i^T\}_{i=1}^n$ , we have by Theorem 2.1 that  $\lambda(\mathcal{A}, p) = p^T M p$ . Thus by Lemma 4.4 and the trivial bound  $\alpha(G) \in [1, n]$ , it holds that

$$(4.6) \quad \min_{p \in \Delta_n} \lambda(\mathcal{A}, p) \in [n^{-1}, O(\log n)].$$

By Observation 4.5, we may—via binary search—refine the confidence interval in (4.6) to size  $n^{-1}$  using  $O(\log(\frac{O(\log n) - n^{-1}}{n^{-1}})) = O(\log n)$  queries to the assumed decision algorithm on  $c\mathcal{A} = \{cA : A \in \mathcal{A}\}$ , for appropriate scalings  $c$ . This yields a polynomial-time<sup>5</sup> approximation to  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  to additive error  $n^{-1}$ , and therefore by Lemma 4.4 yields a polynomial-time approximation  $\alpha(G)$  within a multiplicative factor of  $O(\log n)$ . But this is NP-hard by Theorem 4.3.  $\square$

We remark that although the above reduction was for NP-hardness of *deciding the sign* of  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$ , it also demonstrates an NP-hardness barrier for *approximating the value* of  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$ . This is recorded briefly below. Note that we show hardness for approximating  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  to *additive* error (rather than multiplicative) since this implies hardness for multiplicatively approximating the optimal Lyapunov spectral radius  $R(\mathcal{A}, p) = e^{\lambda(\mathcal{A}, p)}$ , which is the relevant rate of the dynamical system (1.2).

THEOREM 4.6. For all  $\varepsilon > 0$ , it is NP-hard to approximate  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  to an additive error of  $n^{-\varepsilon}$  for a collection  $\mathcal{A}$  of  $n$  symmetric rank-one matrices.

*Proof.* We begin in the same way as in the proof of Theorem 4.1: Given an undirected graph  $G$  on  $n$  nodes, construct the matrix  $M$  and vectors  $\{u_i\}_{i=1}^n$ , let  $\mathcal{A} := \{u_i u_i^T\}_{i=1}^n$ , and note by Theorem 2.1 that  $\lambda(\mathcal{A}, p) = p^T M p$ . Now if  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  can be approximated to an additive error of  $n^{-\varepsilon}$  in polynomial time, then by Lemma 4.4 this yields a polynomial-time algorithm that approximates  $\alpha(G)$  to a multiplicative factor of  $n^{1-\varepsilon} O(\log n)$ . But this is NP-hard by Theorem 4.3.  $\square$

**4.2. Polynomial-time approximation scheme for “well-conditioned” sets of rank-one matrices.** Here, we note that there is a polynomial-time approximation scheme (PTAS) for approximating the design problem  $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  in the case where the set of rank-one matrices  $\mathcal{A}$  is “well-conditioned.” (Note that this does not conflict with the hardness results shown above, as explained below.)

The main idea is that since  $\lambda(\mathcal{A}, p)$  is a quadratic form in  $p$  in this rank-one case by Theorem 2.1, we may use the well-known PTAS for minimizing polynomials over the simplex [12, Theorem 1.7]. Applying the PTAS in [12] to this design problem yields, for any *fixed* constant  $\delta > 0$ , a polynomial-time algorithm that, given any

<sup>5</sup>To be precise, this requires checking that the bit complexity of the inputs to the decision algorithm is also polynomial in  $n$ ; however, this is not an issue. Specifically, in light of Observation 4.5 and the original confidence interval (4.6), each scaling  $c$  used in the binary search satisfies  $\log c \in [n^{-1}, O(\log n)]$ . Moreover, since we refine the search interval only to size  $O(n^{-1})$ , it suffices to truncate each  $\log c$  to precision  $O(n^{-1})$ . Together, this implies that the bit complexity of the scaled matrices  $c\mathcal{A}$  is amplified from that of  $\mathcal{A}$  by at most a polynomial factor in  $n$ .

collection  $\mathcal{A}$  of rank-one matrices, outputs a distribution  $\hat{p} \in \Delta_n$  satisfying

$$(4.7) \quad \lambda(\mathcal{A}, \hat{p}) - \min_{p \in \Delta_n} \lambda(\mathcal{A}, p) \leq \delta \left( \max_{p \in \Delta_n} \lambda(\mathcal{A}, p) - \min_{p \in \Delta_n} \lambda(\mathcal{A}, p) \right).$$

Note that here “polynomial time” means time that is polynomial in the input size (namely  $n$  and the bit complexity of  $\mathcal{A}$ ), but *not* in  $\delta$  (see the discussion in [12, section 1.4]).

Intuitively, this approximation guarantee (4.7) is good whenever the quantity  $\max_{p \in \Delta_n} \lambda(\mathcal{A}, p) - \min_{p \in \Delta_n} \lambda(\mathcal{A}, p)$  governing the error bound is not very large. One concrete setup where this provably occurs is when the set  $\mathcal{A} = \{u_i u_i^T\}_{i=1}^n$  is “well-conditioned” in the sense that all pairwise angles  $\angle(u_i, u_j)$  are uniformly bounded away from  $\frac{\pi}{2}$ —or, in other words, all pairs of  $u_i$  and  $u_j$  are “uniformly away” from being orthogonal. This is made precise in the following result. For simplicity of exposition, we assume that the matrices in  $\mathcal{A}$  are symmetric (i.e., of the form  $u_i u_i^T$ ) and normalized ( $\|u_i\|_2 = 1$ ), but these assumptions can be easily removed.

**THEOREM 4.7.** *Fix any  $\delta, \gamma > 0$ . There is an algorithm that, given any  $\mathcal{A} = \{u_i u_i^T\}_{i=1}^n$  such that each  $\|u_i\|_2 = 1$  and  $|\cos \angle(u_i, u_j)| \geq \gamma$ , outputs  $\hat{p} \in \Delta_n$  satisfying*

$$(4.8) \quad \lambda(\mathcal{A}, \hat{p}) \leq \min_{p \in \Delta} \lambda(\mathcal{A}, p) + \delta \log \frac{1}{\gamma}$$

*in time that is polynomial in  $n$  and the input size of  $\mathcal{A}$  (but not necessarily in  $\delta$  or  $\gamma$ ).*

We make use of the following simple bounds.

**LEMMA 4.8.** *Let  $\mathcal{A} = \{u_i u_i^T\}_{i=1}^n$ . Then*

- $\max_{p \in \Delta_n} \lambda(\mathcal{A}, p) = 2 \max_i \log \|u_i\|_2$ ;
- $\min_{p \in \Delta_n} \lambda(\mathcal{A}, p) \geq \min_{i,j} \log |u_i^T u_j|$ .

*Proof.* By Theorem 2.1,  $\lambda(\mathcal{A}, p) = p^T M_{\mathcal{A}} p$  where  $(M_{\mathcal{A}})_{ij} = \log |u_i^T u_j|$ . For the maximization problem, note that  $2 \max_i \log \|u_i\|_2$  is achieved by  $p = e_i$ . Moreover, this value is optimal since for any  $p \in \Delta_n$ ,

$$\begin{aligned} \lambda(\mathcal{A}, p) &= p^T M_{\mathcal{A}} p = \sum_{ij} p_i p_j (M_{\mathcal{A}})_{ij} \leq \left( \sum_{ij} p_i p_j \right) \max_{ij} (M_{\mathcal{A}})_{ij} \\ &= \max_{ij} (M_{\mathcal{A}})_{ij} \leq \max_i (M_{\mathcal{A}})_{ii}, \end{aligned}$$

where the final inequality is by the definition of  $M_{\mathcal{A}}$  and Cauchy–Schwarz. The bound on the minimization problem follows similarly: For any  $p \in \Delta_n$ ,

$$\lambda(\mathcal{A}, p) = p^T M_{\mathcal{A}} p = \sum_{ij} p_i p_j (M_{\mathcal{A}})_{ij} \geq \left( \sum_{ij} p_i p_j \right) \min_{ij} (M_{\mathcal{A}})_{ij} = \min_{ij} (M_{\mathcal{A}})_{ij}. \quad \square$$

*Proof of Theorem 4.7.* By Lemma 4.8, we have  $\max_{p \in \Delta_n} \lambda(\mathcal{A}, p) - \min_{p \in \Delta_n} \lambda(\mathcal{A}, p) \leq \max_i \log (\|u_i\|_2^2) - \min_{ij} \log |u_i^T u_j| \leq \log \frac{1}{\gamma}$ . The proof is complete by (4.7).  $\square$

Note that the PTAS in Theorem 4.7 is not at odds with the NP-hardness results proved in Theorems 4.1 and 4.6. This is because the runtime is *not* polynomial in the accuracy parameter  $\delta$ , and to obtain the approximation bounds discussed in the hardness results,  $\delta$  must be taken to be scaling with  $n$ —resulting in a runtime that is not polynomial in  $n$ .

**5. Extension to ergodic stationary processes.** So far in this paper, we have considered the analysis and design problems for the Lyapunov exponent in the standard setup where the random matrices are i.i.d. from a discrete distribution. In this section, we discuss extensions to the more general setup of ergodic stationary processes. As corollaries, we discuss setups including i.i.d. processes from continuous distributions, exchangeable processes, and Markov processes. This allows for the study of the stability of more exotic stochastic linear dynamical systems (1.2). The corresponding analysis and design problems are discussed in subsections 5.1 and 5.2, respectively.

**5.1. Ergodic formula for rank-one matrices.** Here we show that, with only a minor modification, the rank-one Lyapunov exponent formula in Theorem 2.1 for i.i.d. processes extends to ergodic stationary processes. The proof is similar, the primary difference being the use of Birkhoff's Ergodic Theorem.

Below, let  $\mathcal{A} = \{uv^T : u, v \in \mathbb{R}^d\}$ , and let it be equipped with the Borel  $\sigma$ -algebra (arising equivalently from any norm on  $\mathbb{R}^{d \times d}$ ). We consider a process  $A_{\sigma_1}, A_{\sigma_2}, \dots$  taking values in  $\mathcal{A}$ . Recall that this process is said to be *stationary* if  $(A_{\sigma_1}, \dots, A_{\sigma_k})$  has the same distribution as  $(A_{\sigma_{\tau+1}}, \dots, A_{\sigma_{\tau+k}})$  for every  $k, \tau \in \mathbb{N}$ . Recall also that this process is said to be *ergodic* if every shift-invariant event has probability 0 or 1. We refer the reader to, e.g., [19, Chapter 7] for further background on such processes.

**THEOREM 5.1.** *Let  $A_{\sigma_1}, A_{\sigma_2}, \dots$  be an ergodic stationary process taking values in  $\mathcal{A}$ , and let  $P$  denote the marginal distribution of  $(A_{\sigma_1}, A_{\sigma_2})$ . If the integrability conditions  $\mathbb{E}|\log \|A_{\sigma_1}\|| < \infty$  and  $\int |\log |u^T v'|| dP(uv^T, u'v'^T) < \infty$  are satisfied, then the (random) quantity*

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_1}\|$$

*exists a.s., and moreover converges to*

$$(5.1) \quad \frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_1}\| \rightarrow \int \log |u^T v'| dP(uv^T, u'v'^T)$$

*a.s. and in  $L_1$ .*

Note that the formula (5.1) is independent of the decompositions  $u^T v$  of the rank-one matrices. This follows by a simple calculation that is identical to the one in Remark 2.2.

*Proof.* Note that the theorem statement is independent of the norm  $\|\cdot\|$  by equivalence of finite-dimensional norms; for convenience of computation, let us take it to be the spectral norm. Consider rank-one decompositions  $A_{\sigma_k} = u_{\sigma_k} v_{\sigma_k}^T$  where w.l.o.g. we take  $\|v_{\sigma_k}\|_2 = 1$ . Just as in (2.2), expand the (random) quantity

$$(5.2) \quad \frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\|_2 = \frac{1}{k} \sum_{t=1}^{k-1} \log |u_{\sigma_t}^T v_{\sigma_{t+1}}| + \frac{1}{k} \log \|u_{\sigma_k}\|_2.$$

To deal with the final term  $\frac{1}{k} \log \|u_{\sigma_k}\|_2$ , observe that by stationarity and the first integrability assumption, we have  $\mathbb{E}|\log \|u_{\sigma_k}\|_2| = \mathbb{E}|\log \|u_{\sigma_k}\|_2 + \log \|v_{\sigma_k}\|_2| = \mathbb{E}|\log \|A_{\sigma_k}\|| = \mathbb{E}|\log \|A_{\sigma_1}\|| < \infty$ . Thus the random variable  $\log \|u_{\sigma_k}\|_2$  is bounded in  $L_1$  (and thus also bounded a.s.), and so

$$(5.3) \quad \frac{1}{k} (\log \|u_{\sigma_k}\|_2 + \log \|v_{\sigma_1}\|_2) \rightarrow 0,$$

where the convergence is a.s. and in  $L_1$ . Now by ergodicity, stationarity, and the second integrability assumption, we may apply Birkhoff's Ergodic Theorem (see, e.g., [19, Theorem 7.2.1]) to obtain

$$(5.4) \quad \frac{1}{k} \sum_{t=1}^k \log |u_{\sigma_t}^T v_{\sigma_{t+1}}| \rightarrow \int \log |u^T v'| dP(uv^T, u'v'^T),$$

where the convergence is a.s. and in  $L_1$ . Combining (5.2), (5.3), and (5.4) completes the proof.  $\square$

We remark that the ergodic formula (5.1) in Theorem 5.1 depends on the process only through its *second-order statistics*. This occurs since the matrices are all rank-one. We now discuss several examples of ergodic stationary processes for which Theorem 5.1 applies.

**5.1.1. I.i.d. processes.** A first, simple example of ergodic stationary processes is an i.i.d. process. Specifically, consider the setup where  $A_{\sigma_1}, A_{\sigma_2}, \dots$  are i.i.d. random variables with (the same) distribution  $\mu$  over  $\mathcal{A}$ . Then Theorem 5.1 shows that, assuming some mild integrability conditions, we have the (a.s. and  $L_1$ ) convergence:

$$(5.5) \quad \frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_1}\| \rightarrow \int \int \log |u^T v'| d\mu(uv^T) d\mu(u'v'^T).$$

Note that this holds for continuous distributions  $\mu$ , i.e., distributions with infinite support. In the special case that  $\mu$  has finite support, we recover Theorem 2.1.

*Example 5.2.* Let  $S^{d-1} = \{z \in \mathbb{R}^d : \|z\|_2 = 1\}$  denote the  $d-1$  sphere, for  $d > 1$ , and let  $\text{Unif}(S^{d-1})$  denote the uniform distribution over it (i.e., the  $(d-1)$ -dimensional spherical measure). In this example, we explicitly compute the Lyapunov exponent over “normalized” rank-one matrices  $\mathcal{A} := \{uv^T : u, v \in S^{d-1}\}$  for the following two probability measures:

- $\mu_{\text{sym}}$  is the induced distribution of  $uu^T \in \mathcal{A}$ , where  $u \sim \text{Unif}(S^{d-1})$ .
- $\mu_{\text{asym}}$  is the induced distribution of  $uv^T \in \mathcal{A}$ , where  $u, v \sim \text{Unif}(S^{d-1})$  and are independent.

By Theorem 5.1, the Lyapunov exponent is the same for both  $\mu_{\text{sym}}$  and  $\mu_{\text{asym}}$ , and is equal to

$$(5.6) \quad \lambda(\mathcal{A}, \mu_{\text{sym}}) = \lambda(\mathcal{A}, \mu_{\text{asym}}) = \mathbb{E}_{u, v \sim \text{Unif}(S^{d-1})} \log |u^T v|.$$

To compute this, we first compute the distribution  $\nu$  of  $u^T v$ . By rotational invariance of the spherical measure  $\text{Unif}(S^{d-1})$ ,  $\nu$  is equal to the distribution of  $z_1$  where  $z \sim \text{Unif}(S^{d-1})$ . This distribution is well known to have density [45]

$$(5.7) \quad \frac{d\nu(t)}{dt} = \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d-1}{2})} (1-t^2)^{(d-3)/2} \mathbf{1}_{t \in (-1,1)},$$

where  $\Gamma(\cdot)$  denotes the Gamma function. Using this, the integral on the right-hand side of (5.6)—and thus the Lyapunov exponents for both  $\mu_{\text{sym}}$  and  $\mu_{\text{asym}}$ —can be computed to be

$$(5.8) \quad \begin{aligned} \mathbb{E}_{u, v \sim \text{Unif}(S^{d-1})} \log |u^T v| &= \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d-1}{2})} \int_{-1}^1 (1-t^2)^{(d-3)/2} \log |t| dt \\ &= -\frac{\Psi(\frac{d}{2}) + \gamma + \log 4}{2}, \end{aligned}$$

TABLE 2

The Lyapunov exponents  $\lambda(\mathcal{A}, \mu_{\text{sym}}) = \lambda(\mathcal{A}, \mu_{\text{asym}}) = -\frac{1}{2}(\Psi(\frac{d}{2}) + \gamma + \log 4)$ , as derived in Example 5.2, for dimensions  $d \in \{2, \dots, 8\}$ . Numerical values reported to six significant digits.

$d$	$-\frac{1}{2}(\Psi(\frac{d}{2}) + \gamma + \log 4)$
2	-0.693147
3	-1
4	-1.19315
5	-1.33333
6	-1.44315
7	-1.53333
8	-1.60981

where  $\Psi(\cdot)$  denotes the Digamma function and  $\gamma \approx 0.5772$  denotes the Euler–Mascheroni constant.

The value of (5.8) is reported in Table 2 for small values  $d \in \{2, \dots, 8\}$ . Note that the value for  $d = 3$  is equal to  $-1$ ; this simple value arises because of the well-known fact that the distribution  $\nu$  of the first coordinate of a random vector on the 2-sphere is the uniform distribution over  $[-1, 1]$  (this can be seen directly from (5.7)). Indeed, using this one can easily derive the special case  $\mathbb{E}_{u,v \sim \text{Unif}(S^2)} \log |u^T v| = \frac{1}{2} \int_{-1}^1 \log |z| dz = \int_0^1 \log z dz = (z \log z - z)_0^1 = -1$ .

The asymptotic behavior of (5.8) as the dimension  $d \rightarrow \infty$  is

$$(5.9) \quad \lambda(\mathcal{A}, \mu_{\text{sym}}) = \lambda(\mathcal{A}, \mu_{\text{asym}}) = -\frac{1}{2}(\log d + \gamma + \log 2) + O(d^{-1}),$$

due to the well-known inequality  $\log z - z^{-1} \leq \Psi(z) \leq \log z - (2z)^{-1}$  that holds for all  $z > 0$  [3]. Note that the leading term  $\log(\frac{1}{\sqrt{d}})$  makes sense intuitively since if  $u, v \sim \text{Unif}(S^{d-1})$  are independent, then  $|u^T v|$  concentrates tightly around  $\frac{1}{\sqrt{d}}$ .<sup>6</sup>

**5.1.2. Exchangeable processes.** Although exchangeable processes are *not* ergodic stationary processes,<sup>7</sup> we can understand them in terms of such processes using De Finetti's Theorem [11]. Specifically, consider the setup of an exchangeable process  $A_{\sigma_1}, A_{\sigma_2}, \dots$  taking values in  $\mathcal{A}$ . Recall that (infinite) exchangeability means that for each  $k \in \mathbb{N}$  and each permutation  $\pi$  of  $\{1, \dots, k\}$ ,  $(A_{\sigma_1}, \dots, A_{\sigma_k})$  has the same distribution as  $(A_{\pi(\sigma_1)}, \dots, A_{\pi(\sigma_k)})$ . Now De Finetti's Theorem—or, more precisely, its generalization in [28] to complete, separable metric spaces—states that the law of this process is equal to a mixture  $\nu$  over i.i.d. product distributions over  $\mathcal{A}$ . It therefore follows from Theorem 5.1 that under the same mild integrability conditions,

<sup>6</sup>Here is a simple, standard way to derive this informally. The starting point is that if  $\tilde{u} \sim \mathcal{N}(\mathbf{0}, I)$ , the normal distribution centered at the origin and with identity covariance, then  $\tilde{u}/\|\tilde{u}\|_2 \sim \text{Unif}(S^{d-1})$  by rotational invariance. Thus  $u^T v$  is equal in distribution to  $\tilde{u}^T \tilde{v}/(\|\tilde{u}\|_2 \|\tilde{v}\|_2)$ , where  $u$  and  $v$  are i.i.d.  $\mathcal{N}(\mathbf{0}, I)$ . Now the numerator  $\tilde{u}^T \tilde{v}$  is equal in distribution to  $\tilde{v}_1 \sim \mathcal{N}(0, 1)$  by rotational invariance, and the denominator  $\|\tilde{u}\|_2 \|\tilde{v}\|_2$  concentrates tightly around  $\sqrt{d}$  since  $\|\tilde{u}\|_2^2 = \sum_{i=1}^d \tilde{u}_i^2$  is  $\Theta(\sqrt{d})$  by the Law of Large Numbers. Thus  $|u^T v|$  is approximately equal to  $|\tilde{v}_1|/\sqrt{d}$ , where  $\tilde{v}_1 \sim \mathcal{N}(0, 1)$ , and this concentrates tightly around  $\Theta(1/\sqrt{d})$ . The dropped terms in this discussion are given precisely in the formula (5.8) and asymptotically by (5.9).

<sup>7</sup>A simple example of an exchangeable process that is not ergodic is the process  $X_1, X_2, \dots$ , where all random variables take the same value  $X$ , and  $X$  is a Bernoulli random variable.

we have the (a.s. and  $L_1$ ) convergence of

$$(5.10) \quad \frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_1}\| \rightarrow \int \int \int \log |u^T v'| d\mu(uv^T) d\mu(u'v'^T) d\nu(\mu).$$

**5.1.3. Markov processes.** Another example of an ergodic stationary process is a Markov process over a countable state space (for simplicity) with an irreducible transition kernel (for ergodicity) and with an initial distribution equal to the stationary distribution (for stationarity); see, e.g., [19, Example 7.1.7]. That is, consider the setup where the non-initial matrices  $A_{\sigma_2}, A_{\sigma_3}, \dots$  are drawn from an irreducible Markov kernel  $Q$  over a countable subset of the rank-one matrices  $\mathcal{A}$ , and the initial matrix  $A_{\sigma_1}$  is drawn from the stationary distribution  $\pi_Q$  of the chain. Then Theorem 5.1 guarantees that, under mild integrability assumptions, we have the (a.s. and  $L_1$ ) convergence

$$(5.11) \quad \frac{1}{k} \log \|A_{\sigma_k} \cdots A_{\sigma_1}\| \rightarrow \int \int \log |u^T v'| \pi_Q(uv^T) Q(uv^T, u'v'^T).$$

Note that this recovers Theorem 2.1 since the original Lyapunov exponent setup is the special case where the columns of  $Q$  are constant, i.e.,  $Q = \mathbf{1}p^T$ , and the stationary distribution is  $\pi_Q = p$ . We also remark that (5.11) still holds when the assumption that the initial matrix  $A_{\sigma_1}$  has stationary distribution  $\pi_Q$  is relaxed.<sup>8</sup>

**5.2. Optimizing the Lyapunov exponent.** We now turn to the design problem of optimizing the convergence rate of the stochastic linear dynamical system (1.2) over different types of stochastic processes  $A_{\sigma_1}, A_{\sigma_2}, \dots$ . We already showed in subsection 4.1 that this design problem is NP-hard for i.i.d. processes. Below, we show that the design problem is also NP-hard for exchangeable processes (subsection 5.2.1) yet admits a polynomial-time algorithm for Markov processes (subsection 5.2.2).

**5.2.1. Exchangeable processes.** Here, we show that the design problem is NP-hard over exchangeable processes.

**THEOREM 5.3.** *Given a collection  $\mathcal{A}$  of  $n$  symmetric rank-one matrices, it is NP-hard to decide whether there exists an exchangeable process  $A_{\sigma_1}, A_{\sigma_2}, \dots$  taking values in  $\mathcal{A}$ , for which the stochastic linear dynamical system (1.2) converges (a.s.) for every initial state.*

To give intuition for this result, recall the following three facts: (i) by De Finetti's Theorem, exchangeable processes are mixtures of i.i.d. distributions; (ii) the Lyapunov exponent is a linear functional in this mixing measure (see (5.10)); and (iii) every linear program with a compact feasible set admits an optimal extreme point. Combining these three facts implies that, among exchangeable processes, there is an optimal i.i.d. process. This explains the hardness in Theorem 5.3, since the design problem over i.i.d. processes is NP-hard by Theorem 4.1. We now formalize this intuition.

*Proof of Theorem 5.3.* By (5.10), there exists such a stabilizing exchangeable process if and only if there exists such a stabilizing i.i.d. process. But deciding whether there exists such a stabilizing i.i.d. process is NP-hard by Theorem 4.1.  $\square$

**5.2.2. Markov processes.** Here we show that the design problem is significantly easier when optimizing over irreducible Markov processes: there is a polynomial-time algorithm that not only can detect whether the associated dynamical system (1.2)

<sup>8</sup>This can be proved using an argument nearly identical to that in the proof of Theorem 2.1.



is stabilizable, but also can compute an irreducible Markov process for which the convergence is fastest. This is in stark contrast to the i.i.d. and exchangeable setups, for which even the stabilization decision problem is NP-hard (see Theorems 4.1 and 5.3).

It will be helpful to introduce some notation. We say a square matrix is *irreducible* if for every simultaneous permutation of its rows and columns, the resulting matrix is a block-diagonal matrix with at most one diagonal block, not counting zero diagonal blocks.<sup>9</sup> We say  $Q \in \mathbb{R}_{\geq 0}^{n \times n}$  is an *irreducible Markov transition matrix* on  $n$  states if  $Q$  is irreducible and  $Q\mathbf{1} = \mathbf{1}$ ; let  $\mathcal{Q}_n$  denote the set of all such matrices. Each  $Q \in \mathcal{Q}_n$  has a unique stationary distribution which we will denote by  $\pi_Q$ .

For simplicity, we assume below an exact arithmetic model. Inexact arithmetic is easily handled by computing logarithms to arbitrary precision.

**THEOREM 5.4.** *There is a polynomial-time algorithm that, given any collection of  $n$  rank-one matrices  $\mathcal{A} = \{u_i v_i^T\}_{i=1}^n$ , outputs a minimizer of*

$$(5.12) \quad \min_{Q \in \mathcal{Q}_n} \sum_{i,j=1}^n (\pi_Q)_i Q_{ij} \log |u_i^T v_j|.$$

Note that the objective function in (5.12) is precisely the convergence rate of the corresponding dynamical system (1.2) by the discussion in subsection 5.1.3. It therefore follows immediately that stability of the dynamical system can be detected in polynomial time: Simply run the algorithm in Theorem 5.4 and check whether the optimization problem (5.12) has strictly negative value.

**COROLLARY 5.5.** *There is a polynomial-time algorithm that, given a collection  $\mathcal{A}$  of  $n$  rank-one matrices, decides whether there exists an irreducible Markov process  $A_{\sigma_1}, A_{\sigma_2}, \dots$  taking values in  $\mathcal{A}$  such that the stochastic linear dynamical system (1.2) converges (a.s.) for every initial state.*

We now turn to the proof of Theorem 5.4. We proceed in two steps, informally sketched as follows. First, we observe that the optimization problem (5.12) can be equivalently recast as a linear program over circulations (Lemma 5.6). Then, by a basic result about decomposing circulations into cycles, we observe that this linear program is equivalent to the Minimum-Mean-Cycle problem in a certain associated graph (Lemma 5.7). The proof is then completed by recalling a well-known polynomial-time algorithm for the Minimum-Mean-Cycle problem.

We now formalize step 1. Below, we say  $F \in \mathbb{R}_{\geq 0}^{n \times n}$  is a (normalized) *circulation* of size  $n$  if  $F\mathbf{1} = F^T\mathbf{1}$  and  $\sum_{ij} F_{ij} = 1$ ; let  $\mathcal{F}_n$  denote the set of all such matrices.

**LEMMA 5.6.**  $\min_{Q \in \mathcal{Q}_n} \sum_{ij} (\pi_Q)_i Q_{ij} \log |u_i^T v_j| = \min_{F \in \mathcal{F}_n} \sum_{ij} F_{ij} \log |u_i^T v_j|.$

*Proof.* For any  $Q \in \mathcal{Q}_n$ , the matrix  $F$  with entries  $F_{ij} := (\pi_Q)_i Q_{ij}$  is easily checked to be in  $\mathcal{F}_n$  since  $F\mathbf{1} = F^T\mathbf{1} = \pi_Q$  by definition of  $\pi_Q$  being the stationary distribution for  $Q$ . This proves the inequality “ $\geq$ .” For the other direction “ $\leq$ ,” consider an optimal  $F \in \mathcal{F}_n$  for the optimization problem on the right-hand side. We may assume w.l.o.g. that  $F$  is irreducible since a circulation can always be decomposed as a convex combination of irreducible circulations (this follows, e.g., from [5, Exercise 7.14]), and each of these irreducible circulations must also be optimal by linearity of

<sup>9</sup>We note that the standard definition of an irreducible matrix does count zero diagonal blocks; this slight tweak allows us to more easily consider optimizing over subsets of  $\mathcal{A}$  without having to introduce additional notation. A similar remark applies to our definition of irreducible Markov transition matrices.

the cost function  $F \mapsto \sum_{ij} F_{ij} \log |u_i^T v_j|$ . Let  $\pi := F\mathbf{1} = F^T\mathbf{1}$ , let  $j^*$  be any fixed index in  $\{1, \dots, n\}$  such that  $\pi_{j^*} > 0$  (such an index exists since  $F \in \mathcal{F}_n$ ), and let  $Q$  be the matrix with entries

$$(5.13) \quad Q_{ij} := \begin{cases} F_{ij}/\pi_i, & \pi_i > 0, \\ 1, & \pi_i = 0, j = j^*, \\ 0, & \pi_i = 0, j \neq j^*. \end{cases}$$

It is straightforward to check that  $Q \in \mathcal{Q}_n$ , that it has stationary distribution  $\pi_Q = \pi$ , and that  $(\pi_Q)_i Q_{ij} = F_{ij}$  for all  $i, j$ . This completes the proof.  $\square$

We now formalize step 2. Below,  $G_{\mathcal{A}}$  denotes the complete directed graph on  $n$  nodes, one for each matrix  $A_i$ , with weight  $\log |u_i^T v_j|$  on the edge from node  $i$  to node  $j$ . The *mean weight* of a cycle in  $G_{\mathcal{A}}$  is the sum of the weights of the edges in the cycle, divided by the total number of edges. The *minimum-cycle-mean* of  $G_{\mathcal{A}}$  is the minimum mean weight of a cycle, where the minimum is over all cycles in  $G_{\mathcal{A}}$ .

LEMMA 5.7.  $\min_{F \in \mathcal{F}_n} \sum_{ij} F_{ij} \log |u_i^T v_j|$  is equal to the minimum-cycle-mean of  $G_{\mathcal{A}}$ .

*Proof.* Recall that a circulation (flow) can be decomposed as the convex combination of cycle (flows); see, e.g., [5, Exercise 7.14]. Thus by linearity of the objective function  $F \mapsto \sum_{ij} F_{ij} \log |u_i^T v_j|$ , there exists an optimal circulation  $F \in \mathcal{F}_n$  and a cycle  $C$  such that  $F_{ij}$  is equal to  $1/|C|$  if  $(i, j)$  is an edge in  $C$ , and 0 otherwise. The direction “ $\geq$ ” then follows since  $\sum_{ij} F_{ij} \log |u_i^T v_j|$  is equal to the mean weight of  $C$ . The other direction “ $\leq$ ” follows by constructing  $F \in \mathcal{F}_n$  from an optimal cycle  $C$  in the same way.  $\square$

We are now ready to conclude the proof of Theorem 5.4.

*Proof of Theorem 5.4.* By Lemmas 5.6 and 5.7,  $\min_{Q \in \mathcal{Q}_n} \sum_{ij} (\pi_Q)_i Q_{ij} \log |u_i^T v_j|$  is equal to the minimum-cycle-mean of  $G_{\mathcal{A}}$ . An optimal such cycle (i.e., one with minimum mean weight) can be computed in  $O(n^3)$  time [31]. Let  $F$  denote the  $n \times n$  matrix with entries  $F_{ij}$  equal to  $1/|C|$  if the cycle contains the edge  $(i, j)$ , and equal to 0 otherwise. Clearly  $F \in \mathcal{F}_n$  and  $\sum_{ij} F_{ij} \log |u_i^T v_j|$  is equal to the mean weight of  $C$ . Now construct  $Q \in \mathcal{Q}_n$  from  $F$  by (5.13) so that  $\sum_{ij} F_{ij} \log |u_i^T v_j| = \sum_{ij} (\pi_Q)_i Q_{ij} \log |u_i^T v_j|$ . Thus  $Q$  is an optimal solution for (5.12). Since  $F$  and  $Q$  are clearly computable from  $C$  in polynomial time, this completes the proof.  $\square$

Remark 5.8. It is known that the Joint Spectral Radius of a collection of rank-one matrices can be computed from the minimum-cycle-mean of  $G_{\mathcal{A}}$  [2, Theorem 2.2]. We have shown here that the analogous optimization problem over irreducible Markov processes is also equivalent to the minimum-cycle-mean of  $G_{\mathcal{A}}$ .

## REFERENCES

- [1] A. A. AHMADI, R. M. JUNGERS, P. A. PARRILO, AND M. ROOZBEHANI, *Joint spectral radius and path-complete graph Lyapunov functions*, SIAM J. Control Optim., 52 (2014), pp. 687–717, <https://doi.org/10.1137/110855272>.
- [2] A. A. AHMADI AND P. A. PARRILO, *Joint spectral radius of rank one matrices and the maximum cycle mean problem*, in Proceedings of the 51st IEEE Conference on Decision and Control (CDC), 2012, pp. 731–733.
- [3] H. ALZER, *On some inequalities for the gamma and psi functions*, Math. Comp., 66 (1997), pp. 373–389.
- [4] Y. BENOIST AND J.-F. QUINT, *Random Walks on Reductive Groups*, Springer, 2016.

- [5] D. BERTSIMAS AND J. N. TSITSIKLIS, *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA, 1997.
- [6] W.-J. BEYN AND A. LUST, *A hybrid method for computing Lyapunov exponents*, Numer. Math., 113 (2009), pp. 357–375.
- [7] V. D. BLONDEL AND J. N. TSITSIKLIS, *The boundedness of all products of a pair of matrices is undecidable*, Systems Control Lett., 41 (2000), pp. 135–140.
- [8] S. BOYD, A. GHOSH, B. PRABHAKAR, AND D. SHAH, *Randomized gossip algorithms*, IEEE Trans. Inform. Theory, 52 (2006), pp. 2508–2530.
- [9] J. E. COHEN, *Subadditivity, generalized products of random matrices and operations research*, SIAM Rev., 30 (1988), pp. 69–86, <https://doi.org/10.1137/1030002>.
- [10] J. E. COHEN, H. KESTEN, AND C. M. NEWMAN, *Random Matrices and Their Applications*, Contemp. Math. 50, American Mathematical Society, 1986.
- [11] B. DE FINETTI, *Sur la condition d'équivalence partielle*, in Actualités scientifiques et industrielles, No. 739, Hermann & Cie, Paris, 1938, pp. 5–18.
- [12] E. DE KLERK, M. LAURENT, AND P. A. PARRILO, *A PTAS for the minimization of polynomials of fixed degree over the simplex*, Theoret. Comput. Sci., 361 (2006), pp. 210–225.
- [13] B. DERRIDA, K. MECHERI, AND J. PICHARD, *Lyapounov exponents of products of random matrices: Weak disorder expansion. Application to localisation*, J. Phys. France, 48 (1987), pp. 733–740.
- [14] J. DEUTSCH AND G. PALADIN, *Product of random matrices in a microcanonical ensemble*, Phys. Rev. Lett., 62 (1989), pp. 695–699.
- [15] M. M. DEZA AND E. DEZA, *Encyclopedia of Distances*, Springer, 2009.
- [16] P. DIACONIS AND D. FREEDMAN, *Iterated random functions*, SIAM Rev., 41 (1999), pp. 45–76, <https://doi.org/10.1137/S0036144598338446>.
- [17] L. DIECI AND E. S. VAN VLECK, *Lyapunov spectral intervals: Theory and computation*, SIAM J. Numer. Anal., 40 (2002), pp. 516–542, <https://doi.org/10.1137/S0036142901392304>.
- [18] L. DIECI AND E. S. VAN VLECK, *Perturbation theory for approximation of Lyapunov exponents by QR methods*, J. Dynam. Differential Equations, 18 (2006), pp. 815–840.
- [19] R. DURRETT, *Probability: Theory and Examples*, Cambridge University Press, 2010.
- [20] R. L. DYKSTRA, *An algorithm for restricted least squares regression*, J. Amer. Statist. Assoc., 78 (1983), pp. 837–842.
- [21] A. FURMAN, *Random walks on groups and random transformations*, in Handbook of Dynamical Systems, Vol. 1A, North-Holland, Amsterdam, 2002, pp. 931–1014.
- [22] H. FURSTENBERG, *Noncommuting random products*, Trans. Amer. Math. Soc., 108 (1963), pp. 377–428.
- [23] H. FURSTENBERG AND H. KESTEN, *Products of random matrices*, Ann. Math. Statist., 31 (1960), pp. 457–469.
- [24] R. GHARAVI AND V. ANANTHARAM, *An upper bound for the largest Lyapunov exponent of a Markovian product of nonnegative matrices*, Theoret. Comput. Sci., 332 (2005), pp. 543–557.
- [25] G. GRIMMETT AND D. STIRZAKER, *Probability and Random Processes*, Oxford University Press, 2001.
- [26] N. GUGLIELMI AND V. PROTASOV, *Exact computation of joint spectral characteristics of linear operators*, Found. Comput. Math., 13 (2013), pp. 37–97.
- [27] L. GURVITS AND A. SAMORODNITSKY, *A note on common quadratic Lyapunov functions for linear inclusions: Exact results and open problems*, in Proceedings of the 44th IEEE Conference on Decision and Control (CDC), 2005, pp. 2350–2355.
- [28] E. HEWITT AND L. J. SAVAGE, *Symmetric measures on Cartesian products*, Trans. Amer. Math. Soc., 80 (1955), pp. 470–501.
- [29] R. JUNGERS, *The Joint Spectral Radius: Theory and Applications*, Lect. Notes Control Inf. Sci. 385, Springer-Verlag, Berlin, Heidelberg, 2009.
- [30] R. M. JUNGERS, V. Y. PROTASOV, AND V. D. BLONDEL, *Overlap-free words and spectra of matrices*, Theoret. Comput. Sci., 410 (2009), pp. 3670–3684.
- [31] R. M. KARP, *A characterization of the minimum cycle mean in a digraph*, Discrete Math., 23 (1978), pp. 309–311.
- [32] E. S. KEY, *Lower bounds for the maximal Lyapunov exponent*, J. Theoret. Probab., 3 (1990), pp. 477–488.
- [33] J. LIU AND M. XIAO, *Rank-one characterization of joint spectral radius of finite matrix family*, Linear Algebra Appl., 438 (2013), pp. 3258–3277.
- [34] R. MAINIERI, *Zeta function for the Lyapunov exponent of a product of random matrices*, Phys. Rev. Lett., 68 (1992), pp. 1965–1968.

- [35] I. D. MORRIS, *Rank One Matrices Do Not Contribute to the Failure of the Finiteness Property*, preprint, <https://arxiv.org/abs/1109.4648>, 2011.
- [36] I. D. MORRIS, *Generic properties of the lower spectral radius for some low-rank pairs of matrices*, *Linear Algebra Appl.*, 524 (2017), pp. 35–60.
- [37] T. S. MOTZKIN AND E. G. STRAUS, *Maxima for graphs and a new proof of a theorem of Turán*, *Canad. J. Math.*, 17 (1965), pp. 533–540.
- [38] P. A. PARRILO AND A. JADBABAIE, *Approximation of the joint spectral radius using sum of squares*, *Linear Algebra Appl.*, 428 (2008), pp. 2385–2402.
- [39] V. Y. PROTASOV, *The generalized joint spectral radius. A geometric approach*, *Izv. Math.*, 61 (1997), pp. 995–1030.
- [40] V. Y. PROTASOV, *On the regularity of de Rham curves*, *Izv. Math.*, 68 (2004), pp. 567–606.
- [41] V. Y. PROTASOV AND R. M. JUNGERS, *Convex optimization methods for computing the Lyapunov exponent of matrices*, in *Proceedings of the 2013 European Control Conference (ECC)*, IEEE, 2013, pp. 3191–3196.
- [42] V. Y. PROTASOV, R. M. JUNGERS, AND V. D. BLONDEL, *Joint spectral characteristics of matrices: A conic programming approach*, *SIAM J. Matrix Anal. Appl.*, 31 (2010), pp. 2146–2162, <https://doi.org/10.1137/090759896>.
- [43] I. J. SCHOENBERG, *Remarks to Maurice Fréchet’s article “Sur la définition axiomatique d’une classe d’espace distanciés vectoriellement applicable sur l’espace de Hilbert,”* *Ann. of Math.* (2), 36 (1935), pp. 724–732.
- [44] D. SUTTER, O. FAWZI, AND R. RENNER, *Bounds on Lyapunov Exponents via Entropy Accumulation*, preprint, <https://arxiv.org/abs/1905.03270v3>, 2019.
- [45] G. SZEGŐ, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ. 23, American Mathematical Society, 1939.
- [46] J. N. TSITSIKLIS AND V. D. BLONDEL, *The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate*, *Math. Control Signals Systems*, 10 (1997), pp. 31–40.
- [47] M. VIANA, *Lectures on Lyapunov Exponents*, Cambridge Stud. Adv. Math. 145, Cambridge University Press, 2014.
- [48] D. ZUCKERMAN, *Linear degree extractors and the inapproximability of max clique and chromatic number*, in *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, 2006, pp. 681–690.